Analyzing User-Level Privacy Attack Against Federated Learning

Year: 2020 | Citations: 225 | Authors: Mengkai Song, Zhibo Wang, Zhifei Zhang, Yang Song, Qian Wang

Abstract

Federated learning has emerged as an advanced privacy-preserving learning technique for mobile edge computing, where the model is trained in a decentralized manner by the clients, preventing the server from directly accessing those private data from the clients. This learning mechanism significantly challenges the attack from the server side. Although the state-of-the-art attacking techniques that incorporated the advance of Generative adversarial networks (GANs) could construct class representatives of the global data distribution among all clients, it is still challenging to distinguishably attack a specific client (i.e., user-level privacy leakage), which is a stronger privacy threat to precisely recover the private data from a specific client. To analyze the privacy leakage of federated learning, this paper gives the first attempt to explore user-level privacy leakage by the attack from a malicious server. We propose a framework incorporating GAN with a multi-task discriminator, called multi-task GAN – Auxiliary Identification (mGAN-AI), which simultaneously discriminates category, reality, and client identity of input samples. The novel discrimination on client identity enables the generator to recover user specified private data. Unlike existing works interfering the federated learning process, the proposed method works “invisibly” on the server side. Furthermore, considering the anonymization strategy for mitigating mGAN-AI, we propose a beforehand linkability attack which re-identifies the anonymized updates by associating the client representatives. A novel siamese network fusing the identification and verification models is developed for measuring the similarity of representatives. The experimental results demonstrate the effectiveness of the proposed approaches and the superior to the state-of-the-art.